Article

Technology Clusters Exploration for Patent Portfolio through Patent Abstract Analysis

Gabjo Kim [1], Joonhyuck Lee [2], Dongsik Jang [2] and Sangsung Park [3,*]

[1] Korea Intellectual Property Strategy Agency, Seoul 06132, Korea; kkjjo@kista.re.kr
[2] Department of Industrial Management Engineering, Korea University, Seoul 02841, Korea;
    iguana751@korea.ac.kr (J.L.); jang@korea.ac.kr (D.J.)
[3] Graduate School of Management of Technology, Korea University, Seoul 02841, Korea
[*] Correspondence: hanyul@korea.ac.kr; Tel.: +82-10-8700-0469; Fax: +82-2-3290-3900

Abstract: This study explores technology clusters through patent analysis. The aim of exploring technology clusters is to grasp competitors' levels of sustainable research and development (R&D) and establish a sustainable strategy for entering an industry. To achieve this, we first grouped the patent documents with similar technologies by applying affinity propagation (AP) clustering, which is effective while grouping large amounts of data. Next, in order to define the technology clusters, we adopted the term frequency-inverse document frequency (TF-IDF) weight, which lists the terms in order of importance. We collected the patent data of Korean electric car companies from the United States Patent and Trademark Office (USPTO) to verify our proposed methodology. As a result, our proposed methodology presents more detailed information on the Korean electric car industry than previous studies.

1. Introduction

In recent years, companies and countries have been responding to complicated and rapid changes in technology and are hence, focusing on establishing structured research development systems to strengthen their capacity for future technology forecasts. In addition, companies and countries are required to continuously invest in technology development to support their sustainable competitiveness [1]. Thus, entering an industry and establishing sustainable business strategies, based on a structured strategy of forecasting promising technology areas and monitoring competitors, is an important issue. In the past, qualitative analysis techniques, such as the Delphi technique, were the main method of technology forecasting and facilitated decisions in establishing a product's improvement or the development plan for new products, as well as grasping technology alternatives and the appropriate time to introduce new technology [2,3]. However, such a method has its disadvantages. First, there is the possibility of ignoring potential relationships between the forecasted events and being influenced by the subjective opinions of the experts. Second, it is costly and time consuming. To supplement these disadvantages, the monitoring of competitors and the study of technology forecasting are extensively used. These studies are performed through patent portfolio development by applying a variety of statistic and data mining techniques on patents that contain a detailed explanation of technology. Patent documents include potential information on developed technologies, such as a summary of the technology and information on inventors and applicants. Such patent information, with its explanation of technology and innovative activity, is a very important source in areas of technology development strategy and market transaction [4]. Therefore, utilizing

patent information appropriately is crucial to technology development as well as business management. Exploring technology clusters in similar technology, grouped by technology similarities in patent documents, is an especially important step in discovering new businesses and understanding the competition in the areas where the business operates. Furthermore, this step contributes to identifying and entering new and inadequately explored technology fields in order to expand the diversification of the business. This research explores the appearance of a new area of technology, which is derived from technology fusion in the patent information of companies in specific technology areas. It also studies which technology areas are to be focused on for a business to survive in a continuously growing industry. This research is separated into two sections—the formation of technology clusters through patent documents analysis, and the extraction of core keywords to define the formed technology clusters. By doing so, we suggest a sustainable strategy to create a lead to enter a new technology industry and discover new business opportunities. In order to achieve this sustainable strategy, this paper presents a methodology to form technology clusters through affinity propagation (AP) clustering based on message passing, wherein patent documents exchange similarities by message. AP is appropriate for patent data, which includes extensive text information, because it clusters the data by measuring similarities in a complicated and abundant amount of data. Furthermore, we extract core keywords of technology clusters through the term frequency-inverse document frequency (TF-IDF) weight in the process of defining the formed technology clusters. We explore the technology clusters from Korean companies in the electric car technology field by analyzing their patent applications with the United States Patent and Trademark Office (USPTO) [5]. From this analysis, we propose a portfolio facilitating new technology development and a sustainable strategy to enter a new industry by forming technology clusters for each business in the area of electric car technology.

2. Literature Review

2.1. Patent Research and Analysis

Much patent research and analysis has been conducted to satisfy patent analysis experts, managers of companies, and their various demands [6]. Patent documents include some information about citation, applicant, inventor, International Patent Classification (IPC), and an abstract indicating a detailed description of the invention. Researching and analyzing this information holds great potential to fully understand the technology levels and innovation trends in countries around the world [7]. In this information, citation can be used to identify a value or innovative activity in a company [8,9]. In addition to citation, there is applicant and inventor information, which shows relative technology leadership in certain areas [10]. Moreover, it is possible to visually realize the influence of technology through network analysis by using IPC information [11]. The abstract in the patent document is especially important information since it provides a detailed description of the technology. However, since the abstract is unstructured data that is hard to directly analyze, it is necessary to transform it into structured data using the text mining technique. The keywords thus extracted from the patent contain core technology, product, components, and methodology of the technology [4]. In this study, we apply the text mining technique to analyze the abstract of the patent to explore the technology clusters.

2.2. Strategy of the Exploring Technology Clusters

In order to establish sustainable research and development strategies, it is important to identify potential future technology and grasp the current level of technology development in a rival company by exploring the technology cluster. Exploring technology clusters by using patent information, which has been approached in various ways by many researchers, and text mining techniques is generally executed in two stages.

The first stage is to form technology clusters from patent documents of similar technologies. Lee et al. [12] transformed unstructured patent data into structured data using the text mining technique. This was followed by principle component analysis (PCA), which was adopted to reduce

the high dimensionality of patent data and to map the patent documents. Kim et al. [13], Jun et al. [14] and Jun [11] developed the technology clusters using k-means clustering based on Euclidean distance, and Jun et al. [15] formed technology clusters using k-medoids clustering. However, there are some limitations—PCA is inefficient for patent mapping in terms of visualization; k-means clustering and k-medoids clustering necessitate the determination of the number of clusters beforehand, and the results of k-means clustering are greatly influenced by the value of the initial cluster center.

The next stage is an interpretation of the formed technology clusters. For this stage, there is a method using terms in the clusters and IPC, indicating the technology classification. The technologies are defined after determining keywords that occur together most frequently in the formed clusters [14–16]. Jun and Park [11] also defined the representative technology of the clusters as the most frequent IPC. Kim et al. [13] suggested the patent map of a semantic network based on the terms' frequency and filing data. However, there are some limitations to defining technology because the definition of technology with unfiltered terms or IPC frequency could contain noise and false or unspecific terms. Therefore, it is necessary to have a more improved and objective approach to exploring technology clusters in order to overcome these limitations.

3. Methodology

3.1. Formation of Technology Clusters

We consider the abstract of the patent document, which describes the technology in detail, for grouping the patent documents with similar technologies. However, as the abstract text is unstructured data, it needs to be transformed into structured data through the text mining technique. In this technique, there is a text normalization process that filters out unnecessary data from information retrieval [17]. The process includes removing blanks, punctuation marks and stop-words such as articles; prepositions; and conjunctions, and changing capital letters to lower case. Following this, the well-organized terms are transformed into a patent-term matrix (PTM) which consists of patent documents in rows and terms in columns. However, there is a problem of sparseness as there are too many terms as compared to the number of patent documents, and most elements are composed of 0. Therefore, multidimensional scaling (MDS) is conducted to overcome this sparseness and reduce the dimensionality. MDS is a method of expressing data as dots in two- or three-dimensional spaces after measuring the characteristics of the data and its similarity or dissimilarity [18]. Thus, the objective of MDS is to visually project the distance between data or similarities on low dimensional space. The distance between *n* data is the measured dissimilarity of that data. The detailed process is as follows: first, a distance matrix $D^{(X)}$, which is the $n \times n$ symmetric matrix of pair-wise dissimilarities, is extracted from the PTM, having n documents and m terms. $D^{(X)}$ having the pair-wise dissimilarities is based on Euclidean distance. When $D^{(X)}$ is given, *n* patent documents are compressed in a lower *p* dimensional space, which is an $n \times p$ configuration matrix. For this, the distance matrix is converted to Gram matrix, as follows:

$$K = -\frac{1}{2}H(D^{(X)})^2 H \tag{1}$$

where, $H = I - (ee^T)/n$, *I* is the identity matrix and *e* is the vector of ones. $K = V^{-1}\Lambda V$ is decomposed using singular value decomposition (SVD), where $\Lambda = [\lambda_1, \cdots, \lambda_n]$ $(\lambda_1 > \cdots > \lambda_n)$, and $V = [V_1, \cdots, V_n]$ is the matrix of normalized eigenvectors. Then, λ_p, which is the *p*th biggest eigenvalue in matrix *K*, becomes an axis of lower dimensional space. Hence, it is possible to get a patent-dimension configuration matrix (PDCM) with *p*-dimensional space, which is lowered by the *i*th component v_i of the *p*th eigenvalue, as shown in Figure 1. Therefore, patent documents are mapped into two- or three-dimensional space. Although three-dimensional space can include more information than two-dimensional space, it leads to a misunderstanding of the internal features of technologies or to a vague relationship between technologies because of dimensional complexity [19]. In this study,

to cluster patent documents with similar technologies, we map patent documents in two-dimensional space, which grasps technology relationships clearly.
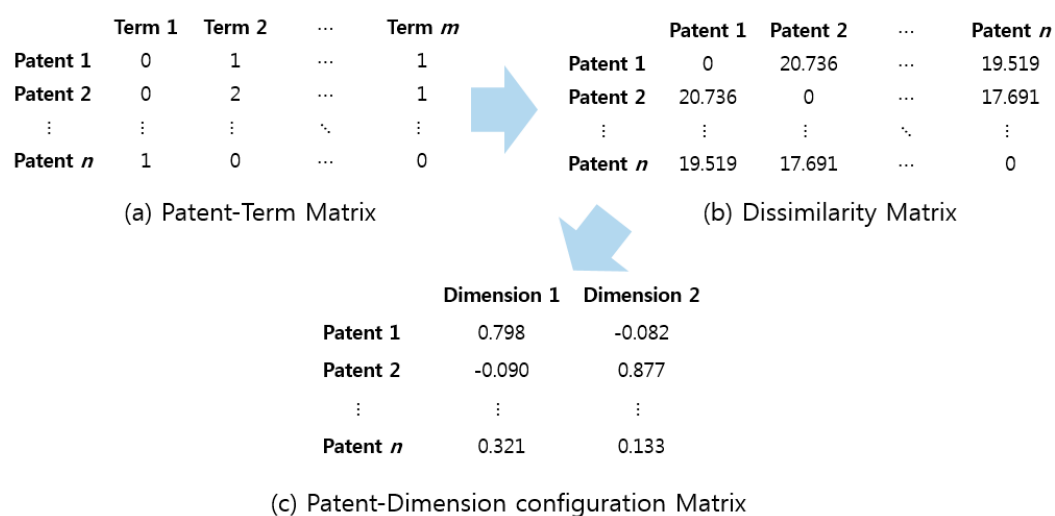
	Term 1	Term 2	⋯	Term *m*
Patent 1	0	1	⋯	1
Patent 2	0	2	⋯	1
⋮	⋮	⋮	⋱	⋮
Patent *n*	1	0	⋯	0

(a) Patent-Term Matrix

	Patent 1	Patent 2	⋯	Patent *n*
Patent 1	0	20.736	⋯	19.519
Patent 2	20.736	0	⋯	17.691
⋮	⋮	⋮	⋱	⋮
Patent *n*	19.519	17.691	⋯	0

(b) Dissimilarity Matrix

	Dimension 1	Dimension 2
Patent 1	0.798	-0.082
Patent 2	-0.090	0.877
⋮	⋮	⋮
Patent *n*	0.321	0.133

(c) Patent-Dimension configuration Matrix

Figure 1. Process of Transforming the PTM into PDCM.

Next, we group the patent documents with similar technology content using AP clustering, which is efficient for big data clustering and does not have to select the initial values of each cluster. AP clustering is a technique based on message passing, which refers to exchanging messages that reflect the similarity of the data points [20]. Exemplars, representing data points, are chosen for each point by exchanging messages. Then each cluster is set by the exemplar. Every data point can be a candidate of exemplar. The AP clustering process proceeds, as shown in Figure 2. First, the similarity between data point *i* and *k* is extracted as an input value. As similarity $s(i, k)$ is an index for showing how data point *k* is appropriate for being an exemplar of data point *i*, Euclidean distance is used to measure the similarity. Instead of choosing the number of clusters beforehand, AP shows a tendency to be chosen as an exemplar by $s(k, k)$ for each data point *k*. $s(k, k)$ is the preference, and the number of exemplars are influenced by the preferences. AP is operated by the processes of the data point working as a network node, and repeatedly exchanging messages from node to node. There are two kinds of messages: responsibility and availability. Responsibility $r(i, k)$ is a message sent from data point *i* to *k*, which indicates how data point *k* is a proper candidate of exemplar for data point *i*. Availability $a(i, k)$ is a message sent from the data point *k* to *i*, indicating a possibility that data point *k* is selected as an exemplar of data point *i*. *lam* is a damping factor. When responsibilities and availabilities are updated, oscillations arise that cause overshooting. To solve this problem, a damping factor is assigned to the old message, which updates the new message so that oscillation can be removed. This process is repeated until the message changes below a threshold or prearranged repeat count. As a result, the value of *k* that maximizes the sum of availability and responsibility is an exemplar of data *i*.

3.2. Definition of Technology Clusters

In order to define the technology in the technology clusters, we apply TF-IDF weight to extract core keywords from the PTM of the technology cluster. TF-IDF weight is a numerical statistic that indicates an ordered importance of terms in the fields of information retrieval and text mining [21]. Term frequency (TF) indicates the number of times each term appears in each document. Inverse document frequency (IDF) is a numerical value that shows unique terms by measuring how many times a certain term occurs in multiple documents, as follows:

$$\text{IDF} = log\frac{N}{n_m} \tag{2}$$

where, N represents the total number of documents and n_m nm indicates the total number of documents including the term m. As the IDF value rises, the term is considered unique in those documents. Therefore, TF-IDF weight is calculated by multiplying TF by IDF. Those with high TF-IDF weight are used as core keywords for defining the technology cluster. The overall proposed methodology is shown in Figure 3.
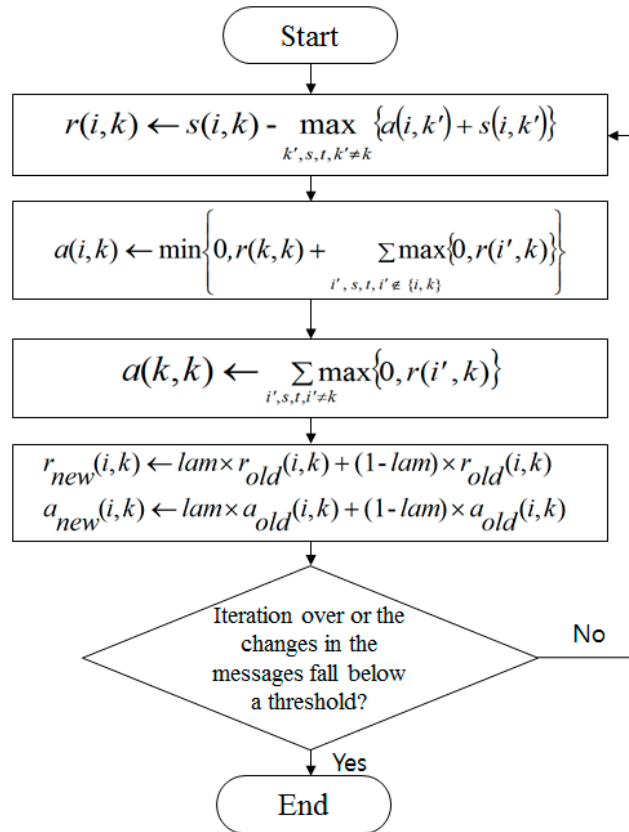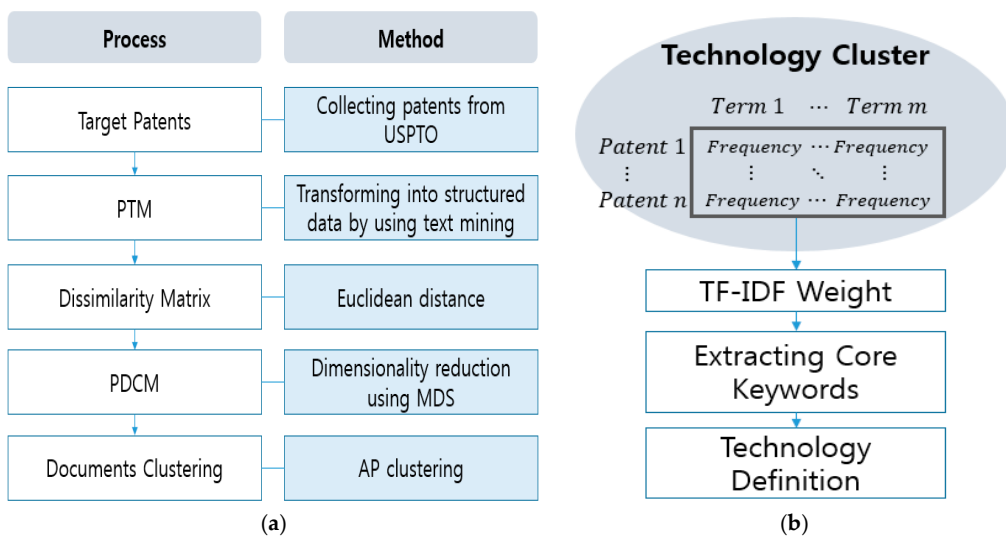

Figure 2. Flow chart of AP clustering.

Figure 3. Overall process of proposed methodology. (**a**) Technology clusters formation; (**b**) Technology Definition.

4. Case Study: Korean Electronic Car Fields

To explore the technology clusters, we collected data of the electric car technology in a Korean company from USPTO. We assumed a company having more than ten patents in the United States as a leading company in electric vehicle technology and analyzed it.

In total, 655 patent documents were collected. Among these documents, Hyundai, Kia, LG, Mando, Samsung, LSIS, Halla, and SK held more than ten patents, as shown in Figure 4. The period of employed patent documents was from 2001 to 2013. We applied the text mining technique to the abstract of the patent document to form the technology clusters for each of the eight companies. Hyundai's dimension of the PTM was 273×2905, which consisted of 273 documents and 2905 terms. Other companies' dimensions were 62×1411 for Kia, 38×928 for LG, 31×798 for Mando, 20×538 for Samsung, 15×303 for LSIS, 13×386 for Halla, and 11×289 for SK. As shown in the Figure 1a above, the PTM of all companies is sparse because it is in high dimensionality and is largely composed of 0. Therefore, we applied MDS in order to overcome this sparseness problem. To achieve this, dissimilarity based on Euclidean distance was calculated, as shown in Figure 1b, and then high dimensionality of terms was reduced to two-dimensionality, as shown in Figure 1c. Next, AP clustering was conducted to form the technology clusters, after mapping the patent documents in two-dimensional space. There are two important parameters in AP clustering—input preference and *lam*. Input preference is correlated with the number of clusters; the number is determined by input preference. Generally, a successful choice for input preference is known to be a median of all the similarities among the data points [20]. Therefore, in this study, we adopted an input preference value according to this theory. *lam* reduces oscillations; the range of *lam* is [0, 1], and this value can be increased if AP cannot converge because of oscillations. Generally, a high value of *lam* leads to steady convergence [20]. In this study, *lam* was set to 0.9. Figure 5 shows the results of AP clustering after arranging each company's patent documents in two-dimensional spaces by MDS.

In Table 1, net similarity is a value that indicates how well exemplars explain the nearby data. Hence, AP could be an objective function that tends to make it to maximum value. We get optimum net similarity by applying input preference with various ranges of values. As a result, each technology cluster is developed for the companies as follows: 19 clusters for Hyundai, 8 for Kia, 7 for LG, 5 for Mando, 5 for Samsung, 3 for LSIS, 3 for Halla, and 3 for SK. For the next step, it was necessary to create a process for technology definition in the developed technology clusters. As shown in Tables 2–9, the patent documents sets of each technology cluster were transformed to the PTM. The technology definition was then inferred from ten core keywords in high value, which were extracted by TF-IDF weight. We requested experts in the electric car field to clearly define the technology through the extracted core keywords.
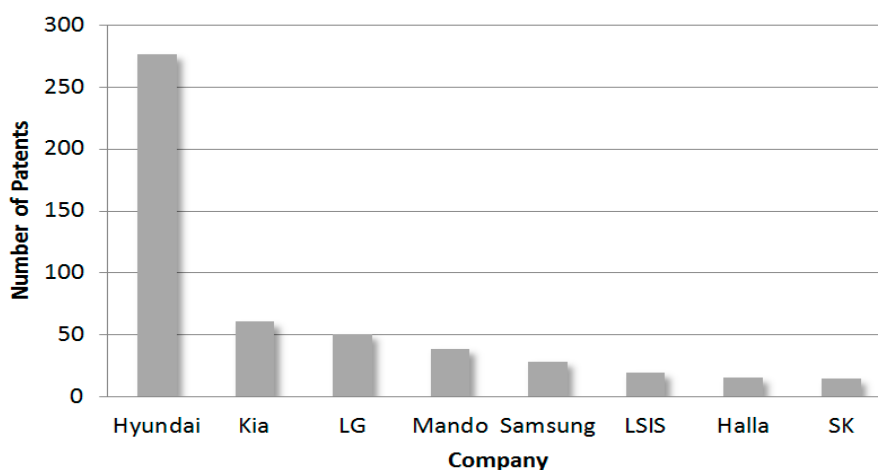

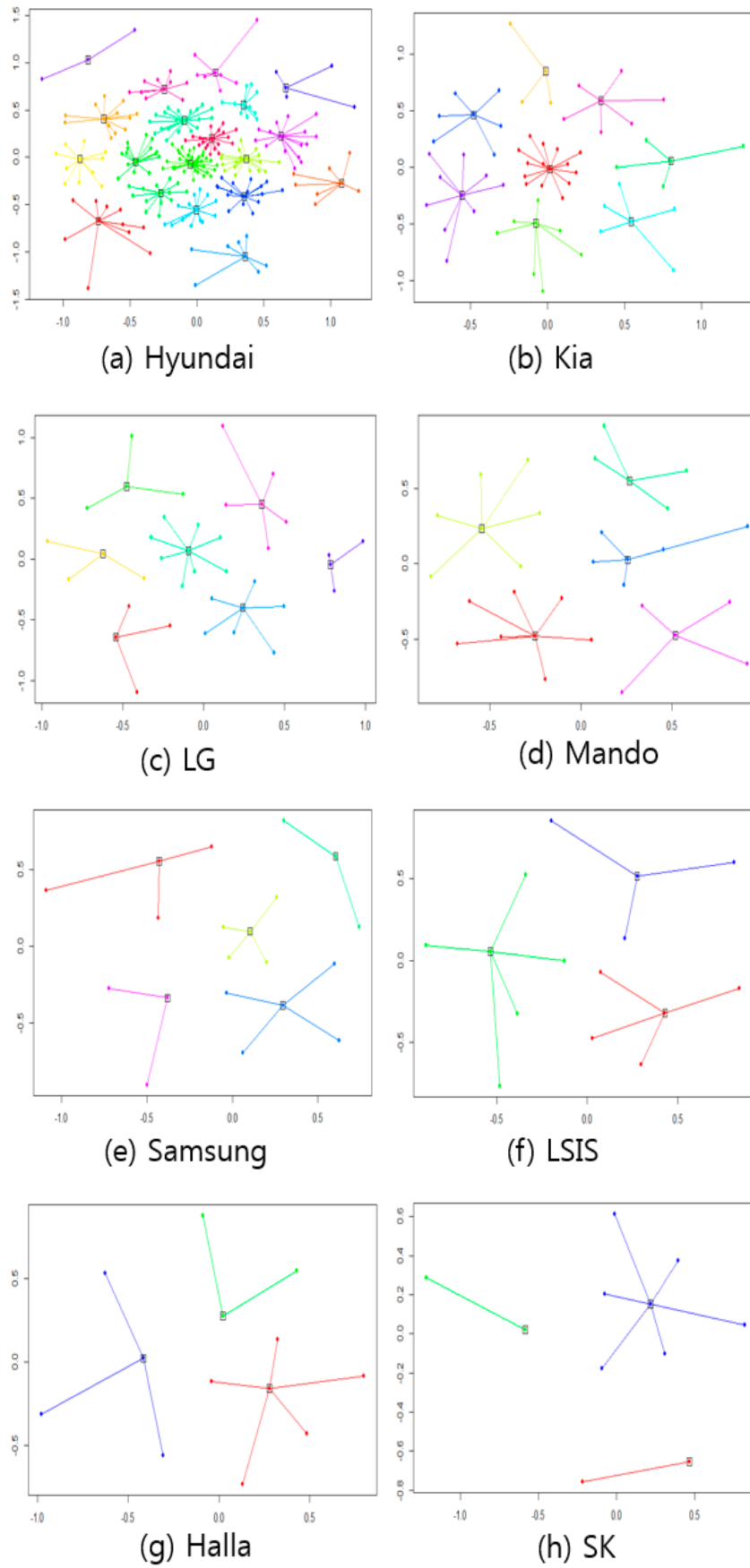**Figure 4.** Collected patent documents of Korean electric car fields from USPTO.

Figure 5. Results of AP clustering of each company in two-dimensional space.

Table 1. Factors of AP clustering of each company's technology.

Company	# of Clusters	Quantity	Net Similarity	Input Preference	# of Iterations
Hyundai	19	273	−22.29912	−0.607961	236
Kia	8	62	−10.03241	−0.635032	189
LG	7	38	−7.79691	−0.684630	139
Mando	5	31	−6.66761	−0.7047505	142
Samsung	5	20	−5.510847	−0.636161	133
LSIS	3	15	−5.002043	−0.710844	123
Halla	3	13	−4.8123	−0.727681	150
SK	3	11	−4.061987	−0.6754607	237

Table 2. Technology clusters for Hyundai.

Cluster	Quantity	Core Keywords	Defined Technology
1	11	three, phase, cam, haptic, set, multifunction, wire, conductive, steering, line	conductive charging
2	8	water, duct, pump, greater, pack, failure, dust, filter, subsequent, motor	water pump control
3	13	measure, PTC (positive temperature coefficient), magnet, limit, maximum, sensor, deceleration, mirror, heat, wheel	PTC
4	11	charge, hydraulic, line, power, tinting, pressure, measure, clutch, base, arm	hydraulic braking
5	18	plasma, grip, part, DTE (distance to empty), ion, noxious, protector, amplifier, coat, differential	plasma reactor for reducing noxious gas
6	38	sun, actuator, conductor, filler, open, slide, terminal, multiple, active, front	multi-zone air conditioning
7	19	heat, oil, temperature, wheel, steer, axle, leg, fashion, stop, speed	protecting a steering system from overheat
8	15	rotor, hydraulic, radiator, cart, wheel, rotary, steer, glass, pressure, drivetrain	cooling system
9	20	virtual, steer, convert, sound, sensor, SOC (state of charge), kick, manipulate, calculate, collect	virtual driving system
10	11	capacitor, link, delay, central, charge, force, pressure, left, right, economic	passive component (capacitor)
11	16	potential, torque, oil, temperature, data, part, type, detect, shaft, change	temperature detecting system
12	10	insert, rotary, aperture, dispose, configure, unit, synchronizer, device, least, provide	synchronize device
13	20	cup, arm, pedal, condenser, occupant, block, window, plate, property, support	vehicle interface devices
14	5	air, compress, force, brake, signal, driven, wheel, pressure, driver, torque	electric compressor (for cooling)
15	3	air, oil, pump, hydraulic, speed, heat, fluid, solenoid, motor, reference	heat pump (for heating)
16	17	high-voltage, recharge, low-voltage, derate, final, regenerative, switch, outlet, consist, remove	regenerative braking
17	8	oil, wiper, pump, available, tire, energy, shift, heat, accumulate, DTE	DTE estimation
18	10	high-voltage, limit, vibrate, oil, signal, bidirectional, grid, smart, unit, brake	smart grid
19	20	regenerator, sensor, coolant, air-condition, remote, house, determine, flow, electron, interior	air conditioning

Table 3. Technology clusters for Kia.

Cluster	Quantity	Core Keywords	Defined Technology
1	15	creep, eco-route, current, door, slide, distance, empty, gradient, driver, manual	eco-routing navigation
2	4	hot, device, heat, configure, fuel, apparatus, first, hybrid, second, separate	heating system for a fuel cell vehicle
3	8	burn, lean, stall, end, apply, drive, energy, desire, HEV (hybrid electric vehicle), position	lean burn drive mode
4	5	count, convert, drive, route, pump, water, lamp, mirror, unit, decouple	control of a converter
5	6	peltier, decelerator, generate, element, exhaust, dispose, leachate, pump, layout, park	heating and cooling system using peltier device
6	7	voltage, high, duct, remain, route, cool, dust, filter, main, amount	high voltage system
7	10	creep, limit, motor/generator, instantaneous, second, capacity, controller, reduction, load, amplifier	creep torque control system
8	7	average, efficiency, past, port, engine, fire, canister, coolant, assemble, cool	temperature control apparatus

Table 4. Technology clusters for LG.

Cluster	Quantity	Core Keywords	Defined Technology
1	4	capacitor, device, relay, control, residual, terminal, strip, switch, measure, configure	relay control apparatus
2	4	cartridge, lead, lithium, connect, middle, film, secondary, accomplished, couple, case	secondary battery
3	4	state, flow, estimate, circuit, connect, charge, configure, module, cell, pre-charge	battery module
4	9	microprocessor, battery, block, interrelation, lead, great, monitor, voltage, calculate, extract	battery monitoring device
5	7	voltage, state, augment, switch, oil, case, make, predict, uncertainty, part	parameter estimation of augmented state
6	4	main, line, modulate, plus, width, condition, differ, energized, fault, feedback	driver circuit and diagnostic method
7	6	model, assemble, state, level, temperature, conduct, dispose, frame, interior, plastic	assembling the battery cell

Table 5. Technology clusters for Mando.

Cluster	Quantity	Core Keywords	Defined Technology
1	7	recovery, gear, automobile, cooperate, step, configure, obtain, velocity, support, adjust	recovery control function
2	6	friction, pulley, light, determine, install, support, coefficient, inside, sense, cumulate	electric power steering
3	5	engine, point, improve, start, time, torque, auxiliary, portion, motor, amount	electronic parking brake
4	6	worm, system, magnet, switch, may, planetary, device, radiation, armature, contact	column-type electric power steering
5	7	coolant, mount, wheel, base, cylinder, hole, passage, plate, solenoid, valve	hydraulic power steering

Table 6. Technology clusters for Samsung.

Cluster	Quantity	Core Keywords	Defined Technology
1	4	turn, rotor, wind, sensor, stator, couple, driver, brake, motor, module	electric blower
2	3	catalyst, control, layer, operate, battery, state, automate, anode, cathode, module	battery development
3	4	air, fuel, electron, emission, output, rectifier, supply, source, oxidant, extend	fuel cell
4	5	stop, terminal, module, power, auxiliary, integrate, configure, charge, contactor, transistor	integrated charging module device
5	5	process, pack, unit, lithium, charge, discharge, current, period, end, consist	lithium battery pack

Table 7. Technology clusters for LSIS.

Cluster	Quantity	Core Keywords	Defined Technology
1	5	communication, program, metadata, first, capacitor, dc-link, power, manage, read, torque	charge management system
2	6	assemble, support, cover, guide, instruct, plate, torque, first, high-voltage, frequency	electric motor control
3	5	signal, communicate, emit, light, drive, shut, bidirectional, end, opposite, case	supplying and blocking of the battery power

Table 8. Technology clusters for Halla.

Cluster	Quantity	Core Keywords	Defined Technology
1	6	Fix, wire, torque, clutch, cell, solar, voltage, eliminate, process, expose	HVAC (heating, ventilation, and air conditioning)
2	3	battery, supply, unit, cool, room, storage, time, duct, display, length	battery cooling
3	4	mode, switch, assemble, detent, load, number, revolution, air-condition, apparatus, value	air conditioner system

Table 9. Technology clusters for SK.

Cluster	Quantity	Core Keywords	Defined Technology
1	6	cell, module, battery, present, use, abnormal, capable, charge, circuit, connection	battery module
2	3	connect, control, dc/dc, convert, module, switch, circuit, SOCi (State of charge based on current), SOCv (state of charge based on voltage), voltage	DC/DC (direct current to direct current) converter
3	4	pre-charge, amount, reactive, cover, circuit, inter-lock, exterior, swell, power, calculate	pre-charge resistance protector

The clustering result shows that Hyundai has 13.92% of multi-zone air conditioning, 7.33% of virtual driving system, 7.33% of vehicle interface devices, and 7.33% of air conditioning as high rate technologies, and 1.10% of heat pump and 1.83% of electric compressor as technologies that are less developed out of a total of 19 technology clusters. Kia is composed of a total of eight technology clusters, in which eco-routing navigation, creep torque control system, and lean burn drive mode technologies are mainly developed, and 6.45% of heating system for a fuel cell vehicle, 8.06% of control of a converter, and 9.68% of heating and cooling system using peltier device technologies are not well-developed. Therefore, it seems that Hyundai and Kia are producing and developing whole components of an electric car. LG's technology clusters comprise a battery monitoring device, parameter estimation of augmented state, and assembling the battery cell as high rate technologies

whereas relay control apparatus, secondary battery, battery model, and driver circuit and diagnostic method are low rate technologies. In the case of Mando, recovery control function, hydraulic power steering, electric power steering technologies, column-type electric power steering, and electronic parking brake technologies are their main technology clusters. Samsung has technology clusters of integrated charging module device and a lithium battery pack as high rate technologies. For LSIS, the technology clusters consist of 31.25% of the charge management system, 37.5% of electric motor control, and 31.25% of supplying and blocking of the battery power. Halla has technology clusters that include 45.15% of heating; ventilation; and air conditioning (HVAC), 23.08% of battery cooling, and 30.77% of the air conditioning system for an electric car. Finally, SK concentrates on battery development, DC/DC converter, and pre-charge resistance protector as their main research and development strategy.

In order to verify the results of our analysis, we compared and analyzed it with the electric car industry and the technology trends report [22,23]. In the report, technology trends and market situations of Korean electric car and electric car components, such as motor and battery, are indicated. In addition, the report is used as a guideline to grasp the technology situation in the Korean electric car industry. For this reason, we used the report to verify the results of our analysis. According to the report, in the past, Hyundai and Kia focused on producing the overall components of the electric car and fuel cell electric car, but have recently begun to focus on the mass production of electric cars. Samsung, LG, and SK are three big players in the battery market for the Korean electric car field; this matched considerably with our text mining results. LSIS focused on the changing infrastructure sector and motor development, and Halla focused on research and development of air conditioning systems. These results correspond to our analysis results. Mando develops the technology of regenerative brakes and chargers. Hence, we were able to find out that the results of our analysis closely matched the report; moreover, our results show the technology development status of each company in more detail. However, there were a few differences between the results of the patent analysis and the industry trends report, as patents usually take years to commercialize [24]. In addition, as shown in Table 10, the technology trends reported suggested the concepts of technologies that are mostly applied to the final products, whereas patents mainly specified the concepts of components and elemental technologies. This, thus, implies that the present study can provide an important piece of information in predicting which technologies will be commercialized and in understanding their details.

Table 10. Korean electric car companies' technology trends in the trends report.

Company	Trends Report [22,23]	
	Present Business Market	**Future Business Plan and Technology**
Hyundai	FSEV (full speed electric vehicles) Motor	Electric car manufacturing Electric car charging technology
Kia	FSEV Motor	Electric car manufacturing Electric car charging technology
LG	Battery Renerative brakes	Li-Ion HEV and PHEV (plug-in hybrid electric vehicle) battery Rapid Charging system
Mando	Renerative brakes Charger	Electric car embedded charging system, charger, condenser development
Samsung	FSEV Battery Cell	Electric car charging system for parking lot of residential complex
LSIS	Inverter Motor	Electric car charging stand manufacturing and electric car charging infrastructures construction
Halla	Work Machine Battery Air conditioning System	Battery pack Air conditioning System
SK	Battery Infrastructure	Combined Li-Ion Cell

5. Conclusions

In this paper, we proposed a method to explore technology clusters to develop an entry strategy by developing technology and monitoring competitor technology levels in the Korean electric car industry. To achieve this, this study included two kinds of themes, as follows: first, AP clustering was suggested for developing the technology clusters. In the process of developing the technology clusters, which group the patent documents with similar technologies, unstructured patent documents were transformed into structured data, and then the patent documents were mapped on low dimensional space using MDS. Next, TF-IDF weight was conducted for extracting core keywords in the process of defining the technology clusters. This study used AP clustering and TF-IDF especially to overcome the problems of prior studies, which are selecting the number of clusters in advance and defining the technology clusters based on the frequency of terms, including noise terms. With our case study, we were able to indicate the Korean companies' research situations and the direction of the electric car industry. This study is also expected to provide a patent portfolio, which is sustainable in developing research and development for companies that are about to enter the electric car industry. By exploring technology clusters of each company's patent information, we examined which technologies the companies have, and how these technologies are concentrated and distributed. This is expected to be helpful for improving competitive power and reducing the cost of sustainable research and development for many companies. In particular, compared to previous reports, the approach in this report made it possible to determine the technology trends in more detail, and predict the technologies that are likely to be commercialized in the future. Despite the comprehensive and effective aspects of our proposed method, there are still some limitations. First, AP clustering is sensitive to the number of clusters in accordance with a set of parameters. In addition, there was qualitative judgment from experts over the interpretation of the technology clusters by using extracted core keywords, even though we removed unnecessary terms by using TF-IDF weights. Therefore, future studies should combine more diverse and effective techniques of data mining and quantitative analysis to develop a more advanced and objective patent portfolio.

Author Contributions: Gabjo Kim designed this research and collected the data set for the experiment. Joonhyuck Lee and Dongsik Jang analyzed the data to show the validity of this study. Sangsung Park wrote the paper and performed the entire research steps.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dickinson, M.W.; Thornton, A.C.; Graves, S. Technology portfolio management: Optimizing interdependent projects over multiple time periods. *IEEE Trans. Eng. Mang.* **2001**, *48*, 518–527. [CrossRef]
2. Dalkey, N.; Helmer, O. An experimental application of the Delphi method to the use of experts. *Mang. Sci.* **1963**, *9*, 458–467. [CrossRef]
3. Silberglitt, R.; Anton, P.S.; Howell, D.R.; Wong, A. *The Global Technology Revolution 2020, In-Depth Analyses*; RAND Corporation: Santa Monica, CA, USA, 2006.
4. Son, C.; Suh, Y.; Jeon, J.; Park, Y. Development of a GTM-based patent map for identifying patent vacuums. *Expert Syst. Appl.* **2012**, *39*, 2489–2500. [CrossRef]
5. The United States Patent and Trademark Office. Available online: http://www.uspto.gov (accessed on 16 December 2013).
6. Abbas, A.; Zhang, L.; Khan, S.U. A literature review on the state-of-the-art in patent analysis. *World Pat. Inf.* **2014**, *1*, 1–11. [CrossRef]
7. Ejermo, O. Regional innovation measured by patent data-Does quality matter? *Ind. Innov.* **2009**, *16*, 141–165. [CrossRef]

8. Chen, Y.; Chang, K.C. The relationship between a firm's patent quality and its market value: The case of US pharmaceutical industry. *Technol. Forecast Soc.* **2010**, *77*, 20–33. [CrossRef]

9. Hall, B.H.; Jaffe, A.B.; Trajtenberg, M. Market value and patent citations. *Rand. J. Econ.* **2005**, *36*, 16–38.

10. Ernst, H. Patent information for strategic technology management. *World Pat. Inf.* **2003**, *25*, 233–242. [CrossRef]

11. Jun, S.; Park, S.S. Examining technological innovation of Apple using patent analysis. *Ind. Manag. Data Syst.* **2013**, *113*, 890–907. [CrossRef]

12. Lee, S.; Yoon, B.; Park, Y. An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation* **2009**, *29*, 481–497. [CrossRef]

13. Kim, Y.G.; Suh, J.H.; Park, S.C. Visualization of patent analysis for emerging technology. *Expert Syst. Appl.* **2008**, *34*, 1804–1812. [CrossRef]

14. Jun, S.; Park, S.S.; Jang, D.S. Patent management for technology forecasting: A case study of the bio-industry. *J. Intellect. Prop. Rig.* **2012**, *17*, 539–546.

15. Jun, S.; Park, S.S.; Jang, D.S. Technology forecasting using matrix map and patent clustering. *Ind. Manag. Data Syst.* **2012**, *112*, 786–807. [CrossRef]

16. Choi, S.; Jun, S. Vacant technology forecasting using new Bayesian patent clustering. *Technol. Anal. Strateg.* **2013**, *26*, 241–251. [CrossRef]

17. Toledo-Alvarado, J.I.; Guzmán-Arenas, A.; Martínez-Luna, G.L. Automatic Building of an Ontology from a Corpus of Text Documents Using Data Mining Tools. *J. Appl. Res. Technol.* **2012**, *10*, 398–404.

18. Cox, T.; Cox, M. *Multidimensional Scaling*, 2nd ed.; Chapman Hall: Boca Raton, FL, USA, 2001.

19. Yoon, B. Strategic visualization tools for managing technological information. *Technol. Anal. Strateg.* **2010**, *22*, 377–397. [CrossRef]

20. Frey, B.J.; Dueck, D. Clustering by Passing Messages between Data Points. *Science* **2007**, *315*, 972–976. [CrossRef] [PubMed]

21. Salton, G.; Buckley, C. Term-weighting approaches in automatic text retrieval. *Inform. Process. Manag.* **1998**, *24*, 513–523. [CrossRef]

22. Virtanen, A.; Lee, Y.M. Electric Vehicles South Korea, Finpro, Seoul. Available online: http://www.finpro.fi/documents/10304/9b52fef5-e0d1-4938-8ec8-961ea6c87491 (accessed on 2 April 2014).

23. Anderman, M. Assessing the Future of Hybrid and Electric Vehicles: The xEV Industry Insider Report Based on Private Onsite Interviews with Leading Technologists and Executives. Advanced Automotive Batteries. Available online: http://www.yano.co.jp/pdf/announce/Executive_Summary_Selections.pdf (accessed on 10 October 2016).

24. Chandy, R.; Hopstaken, B.; Narasimhan, O.; Prabhu, J. From invention to innovation: Conversion ability in product development. *J. Mark. Res.* **2006**, *43*, 494–508. [CrossRef]